Ph.D. Dissertation of Engineering

# Predicting Cellular Branching with Inverse Cellular Automata via Recurrent Neural Network

## 재귀 신경망이 적용된 역 셀룰러 오토마타 학습법을 활용한 세포 분지(分枝) 예측

February, 2018

Graduate School of Mechanical Engineering
Seoul National University
Mechanical Engineering Major

Kyungwon Yun

# 재귀 신경망이 적용된 역 셀룰러 오토마타 학습법을 활용한 세포 분지(分枝) 예측

# Predicting Cellular Branching with Inverse Cellular Automata via Recurrent Neural Network

지도교수 전 누 리

이 논문을 공학박사 학위논문으로 제출함

2017 년 12 월


서울대학교 대학원

기계항공공학부

윤 경 원


윤경원의 공학박사 학위논문을 인준함

2017 년 12 월

위 원 장 : 이 건 우

부위원장 : 전 누 리

위 원 : 고 승 환

위 원 : 방 현 우

위 원 : 윤 호 영

# Abstract

Biological processes are very complex so the predicting of such processes is considered to be very difficult. When it comes to the image-based prediction, one of the main reason of the predicting difficulty comes from the cost that gathering enough visual data which is extremely expensive. In this thesis, we introduce a novel method to reduce the cost of the prediction of cellular branching with Inverse Cellular Automata (ICA). This method assumes the biological image sequence as the set of cellular automata rules between *t* and *t+1*. With this assumption, single biological image sequence can provide *width * height * frame* number of cellular automata set of rule samples so they can be used as traditional Recurrent Neural Network(RNN) based machine learning. Proposed method overcomes the problem of lack of samples for machine learning in biological image predicting. We first briefly review the traditional cellular automata and introduce the concept of the ICA. And we will check the learnability with existing automata image rules. Finally, we show the prediction images with proposed method.

Keyword : Cellular Automata, Biological Imaging, Image Prediction, Artificial Neural Network, Recurrent Neural Network
Student Number : 2012-30924

# Table of Contents

# List of Tables

# List of Figures

# Chapter 1.

# Introduction

## 1.1. 2-D biological imaging assay and biological process

In this full thesis, we would like to discuss biological imaging, especially 2-D image taken by optical microscope. Here, the term 'imaging' shortly stands for taking an image or image sequence with a 2-D optical microscope for biological research purpose.

Investigation and discovery of an unknown biological process has been one of the fundamental research goal in biology. And optical microscopy is still playing a critical role for those discoveries. Many methods have been developed both from microscopy and biology to investigate various biological phenomena. In biology, for example, researchers use staining of certain protein, in order to mark them as fiducial indicator of a process. Green fluorescent protein or GFP, is one of the most well-known methods of protein staining.

Carefully taken single image contains plenty of static information, such as shape, size, or number of target organelles, configuration of cells, etc. There are several topics that can be investigated only with repetition of still imaging. However, researchers are not only interested in single moment of interest, but also interested in whole biological process. In this case, researchers required to consider time-based imaging. Live imaging is to get time dependent information, taking continuous image with short time gap that is ignorable enough to investigate research target change. However, live imaging required to place target specimen onto the microscope for a long time, researchers should take account of cellular damage. Also, live imaging forces to monopoly of the microscope for only single specimen, results in dissipate of research resources. Fortunately, biological process has not only short term change that requires live imaging, but also long-term change. Time-lapsed imaging compromises the gap between still imaging and live imaging.

## 1.2. Limitations of Emergent Feature Assay

Biological process itself is self-organization process. Thus, every target specimen for imaging in biological study are inevitably product of highly complex self-organization process. Although numerous methods have been developed to allow researchers conveniently investigate various features of biological interest, the aids that have given to researchers are usually very basic transformation and transportation of information. In other words, many methods for biological observation or investigation, regardless of qualitative or quantitative study, understanding and analysis of complex biological features still mainly depends on human judgment.

There are various methods for studying complex biological behaviors. (Cell counting, branch counting, volume measurement, etc.) However, these feature recognition methods generally depends on model-based algorithm. This means the recognition process premises human understanding of the target feature, and usually defined by a few number of simplified properties (number, size, color, etc.).

Besides, during the simplification, we usually lose information those are contained in the original complex biological process. The more complex the target feature becomes so the feature cannot be defined by a few number of properties, the higher dependency to the human judgment is required. Further, when the target in interest comes to biological process itself, not the stationary features, algorithmic modeling becomes more difficult. It is because the change in biological process itself is another highly complex behavior, and it usually has more complexity than the stationary property.

Human's recognition of the world highly depends on the visual recognition, and our visual recognition is usually restricted in 2-dimensional space. We cannot recognize the time dimension as we do in 2-D image recognition. We can recognize various local diversity in 2-D image simultaneously, but we cannot recognize the various local 'changes' which are existing in time dimension simultaneously as we do in 2-D image. Since the change in time dimension is far beyond of human recognition range, it is usually extremely difficult to model the time-dependent complex changes in biological process. And that is why we do have to find

roundabout and complementary methods to investigate complex biological process such as Fourier-transform or statistics for the time dimension.

## 1.3. Machine Learning and Artificial Neural Network

Since the electronic image capture devices have developed, researchers continuously studied to achieve machine-based high-level understanding of given scene. Area named machine vision or computer vision (CV) thus deeply explored into the problem of image understanding, and successfully solved many tasks such as classification [1], object recognition and identification, object tracking, feature detection, etc [2]. Recently, artificial neural network (ANN) or deep-learning [3, 4] proposes a new spectrum of image understanding, and greatly surpasses many achievements, so we now prospect human level completeness in certain image based tasks.

Although recent achievements of ANN based image processing promises highly reliable image prediction, there are several problems to apply the method directly to the biological image prediction. One of the biggest problem is that in biological imaging, it is generally hard to provide enough image samples to the neural network learning. Despite the recent high-throughput biological experiments can perhaps reproduce hundreds of biological processes, it is generally not enough

to guarantee the high prediction accuracy in artificial neural network learning, which requires more than $10^4$ of samples in general. Thus, to overcome this number of sample problem in biological image prediction, one might repeat the thousands of high-throughput experiments to provide enough samples for machine learning, only for single biological process, which is highly ineffective.

## 1.4. Cellular Automata

Meanwhile, in computer science and mathematics, Cellular Automata theory [5] has been received attention since the theory resembles the biological process, generating highly complex behavior with only small set of rules. The theory first suggested in 1940's, by John von Neumann [6] and Stanislaw Ulam [7], respectively. They first wanted to developed the self-replication system. After several throughout, the theory received attention when the 'Conway's Game of Life' introduced to the academia in 1970's.[8]

The typical cellular automata system divides the given space into same size of cells, which has finite number of states. And the system has rules for cells to follow to decide the next state of the given cell. The rule here is generally given by the set of states of 'neighborhood' cells, which means the adjacent cells of the cell in interest, and the state of the cell itself in interest. (Figure 1.1)

The first characteristic feature of the CA is that in CA system, the information of the target cell is only transferred to its neighborhood cells. The world we model has basically has inborn limitation in terms of information communication. Every

information requires a certain amount of time to transfer the information to a receiver. For example, in order to change the morphology of a cell, certain type of chemical material should be transferred into the cell organelle along to the working pathway, which requires time to material transfer or exchange. And the inborn characteristic of CA receives the information from only the target-cell's surroundings, which limits the information transfer both to space and time. In other words, in CA, target cells status of next generation receives only information of previous status of itself and its surroundings, and other cells do not affect the status of the target cell.

The other feature of CA is the availability of generating chaotic consequences. As revealed by Stephan Wolfram [9], yet simple rule can result in complex results. For example, rules such as 30, 45, 73 in 1-D binary cellular automata shows unpredictable results from the very early stage of the generation. (Figure 1.2.)

Because of these two characteristic features, CA also received attention in biological area. There are number of researches to apply the CA into biological modeling. [10-13](Figure 1.3.)

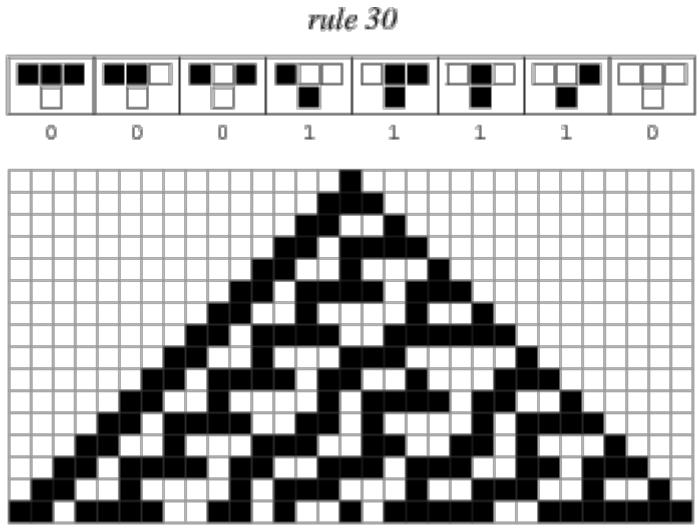


Figure 1.1. Structure of typical Cellular Automata Rules. Picture from [9]

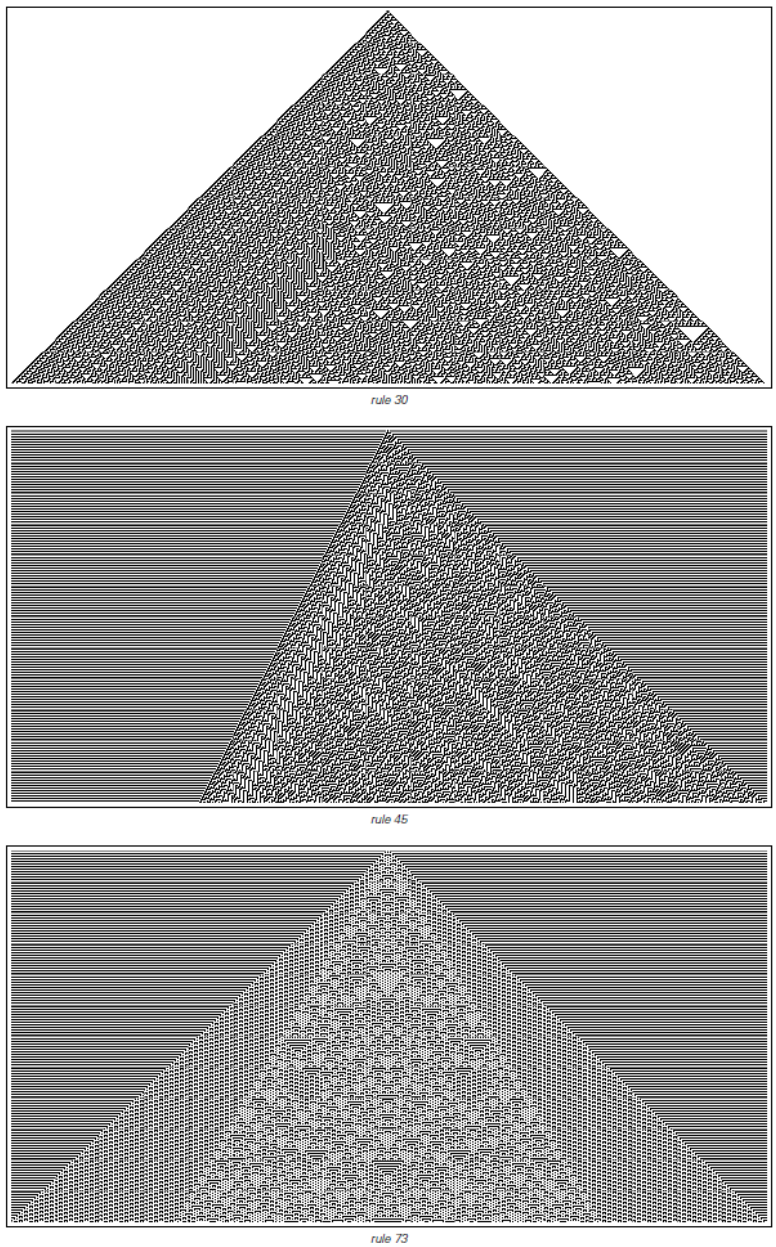


Figure 1.2. Examples of CA rules. [9]

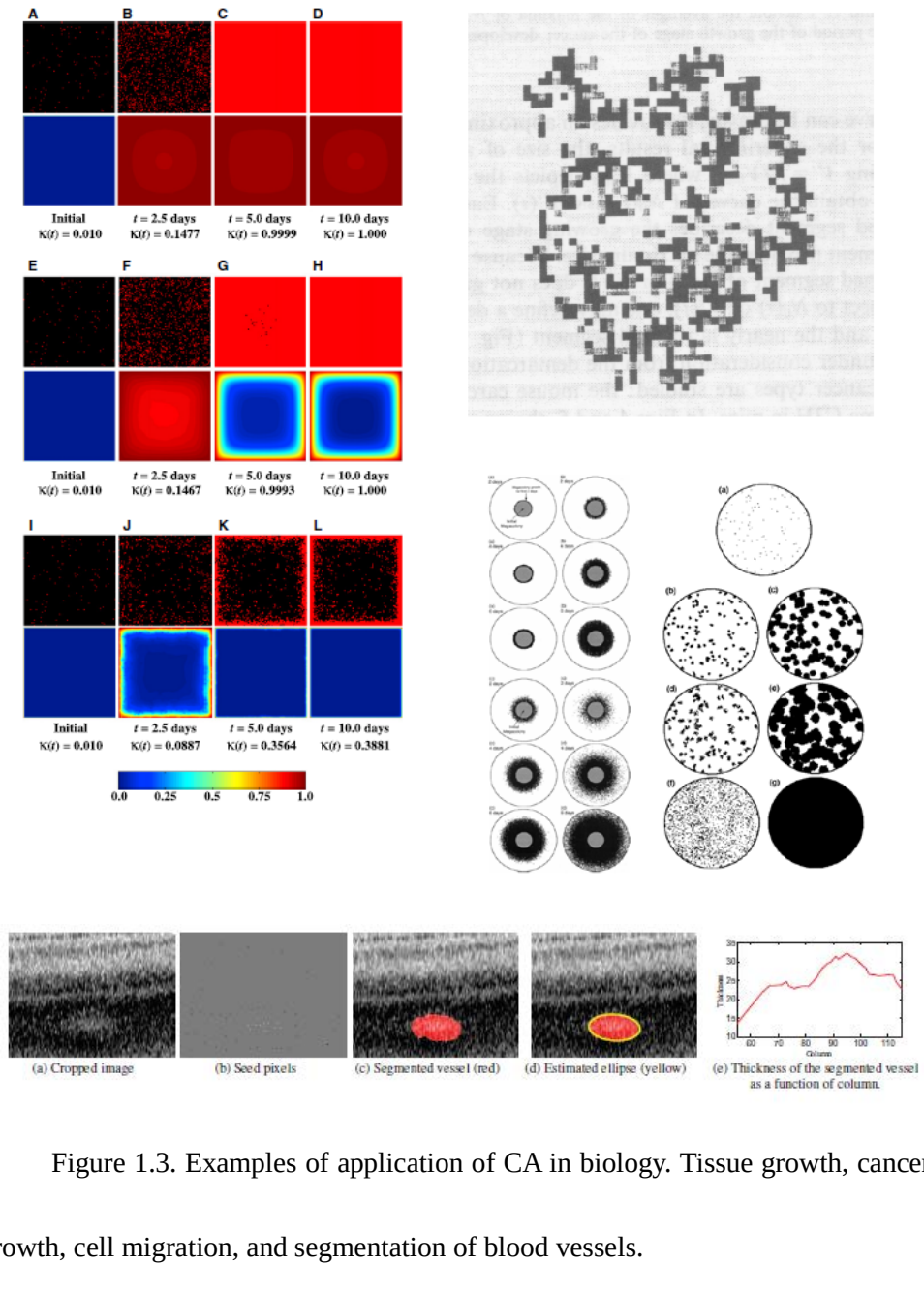


Figure 1.3. Examples of application of CA in biology. Tissue growth, cancer growth, cell migration, and segmentation of blood vessels.

## 1.5. Overview

In this thesis, we propose a novel way to predict the biological process via applying cellular automata into the artificial neural network, to overcome the lack of sample problem in typical neural network based machine learning. In Chapter 2, We first introduce the concept of the inverse cellular automata, which assumes the biological image sequence as the set of automata rules for artificial neural network based machine learning. In Chapter 3, we will check the learnability of the Inverse Cellular Automata via applying proposed method to the existing automata rules such as Conway's Game of Life. In Chapter 4, we will apply the proposed method into the real biological image sequence to predict its behavior. In Chapter 5, we will discuss the result followed by conclusion.

# Chapter 2.

# Inverse Cellular Automata Training

## 2.1. Introduction

In this research, we are interested in the features of cellular automata discussed above. One was that a simple rule can results in chaotic results, and the other was that the spatial information is only affected by its surroundings. Since the biological process is one of the representative chaotic process, and the process is affected by its surroundings, we assumed the short-term time-lapsed microscope biological imaging is also can be treated as cellular automata process. Once we assume the process as cellular automata process, we can deal with the biological live imaging as 2-D cellular automata process.

After we consider the 2-D live imaging of biological process as 2-D cellular automata, we wonder the automata rules for the process. In the typical cellular automata process, the next generation is decided by the known rule. So the process is generally decisive process. Meanwhile, when we have the 2-D live imaging source and the consider the source as cellular automata process, the rule between frames

remains unknown. In this work, we apply machine learning method and learn the unknown rules of cellular automata from the 2-D live imaging. In this thesis, we call the procedure that learns the rules of cellular automata from the biological imaging as ‘Inverse Cellular Automata’ or ICA in short.

In this chapter, we briefly introduce the basic concept of the inverse cellular automata and the methods to achieve the cellular automata.

## 2.2. Methods

### 2.2.1. Overview

The main aim of this chapter is too introduce the method to learn the cellular automata rule from the 2-D live image sequence. In order to apply the cellular automata into the biological image processing, we first need to preprocess the live image as learnable format. As in [Table 2.1.], each frame of original 2-D live image is considered as single generation of cellular automata process. And, the single pixel is considered as the single cell in the cellular automata terms of view. Then the status of the cells becomes the value of each pixel. To make the problem simple, we first only consider about the grayscale images which has single channel for the brightness of the image.

| CELLULAR AUTOMATA | INVERSE CELLULAR AUTOMATA FROM 2-D IMAGING |
|---|---|
| GENERATION | Image Frame |
| CELL | Single Pixel |
| STATUS | Pixel Value |

Table 2.1. Comparison of CA and ICA for 2-D Imaging.

In this sense, we can compare the original cellular automata process and the inverse cellular automata. In the inverse cellular automata, the rule is learned from the next generation's (frame) status of the target cell surroundings. In other words, we train the rule of the CA from the known status. When we originally want to predict the next stage of the biological process, researchers generally consider the whole image of the certain time step to predict the next stage. However, in the ICA, we learn the rules to determine the status of the next generation from the relatively small region of interest, or ROI, not the whole image. Here, we break down the big problem into the small sub-problems to solve the original problem. Once we learn the CA rules properly, we can simply apply the learned rules recursively to the initial frame. In the sense, we can call the ICA as the divide and conquer process that one of the famous problem solving process.

### 2.2.2. Learning Automata Rules

The procedure to learn the automata rule from the bio-imaging follows steps below. First, we need to capture the target of interest cells via optical microscope. The source image requires several conditions. Since the rule learning depends on the physical location of each cells that defines the status of cellular automata, image needs to be captured in constant position. In other words, drafting or difference in lighting condition between frames must affect the final learning performance. And time step between frame should be remain constant too. Second step is to preprocess the captured image sequence to enhance the contrast within a single image. Here, same image preprocess step should be applied to all image sequence to maintain the image lighting constancy. If there are several different image processing methods were applied to different image frames, the rule learning for ICA cannot be stable.

Third step is to reshape the data structure to be learnable way. [Figure] Reshaping dimension depends on the dimension of surrounding area. In practice, we used 3 by 3 window which as 8 surrounding cells to determine the next status of the

target cell. Also, the boundary is padded along to the window size. Specifically, assuming odd window size, the padding size is determined by

$$\text{pad size} = \frac{(\text{window size} - 1)}{2}$$

Then the dimension of single training data becomes,

$$\text{dimension of single training data} = (2 * \text{pad size}) + 1$$

which equals to square of the window size.

And the total number of source training data becomes,

$$\text{total \# of training data} = \text{image width} \times \text{image height} \times (\text{\# of frames} - 1)$$

Here, -1 comes from the last frame which has no next frame and can't have the index.

Forth step is to match the target ROI's status and the center of the ROI in the next frame as sample input and output for machine learning. For the target cell in the n-th frame, the center of ROI in the n+1-th frame becomes the index of the target ROI, which is the status of the next generation of the target cell in general cellular automata terminology.

Fifth step is to decompose the data into sets representing training set, test set, and validation set. For this step, first we randomly mixed the ordered reshaped data for remove the correlation between ordered data. Then the 80% of randomly mixed reshaped data are used as a training set, and the 15% are used as a test set, and 5% are used as a validation set.

Finally, prepared set is used to train the neural network we used. We used the modified recurrent neural network (RNN) model to train the ICA.[14-16]

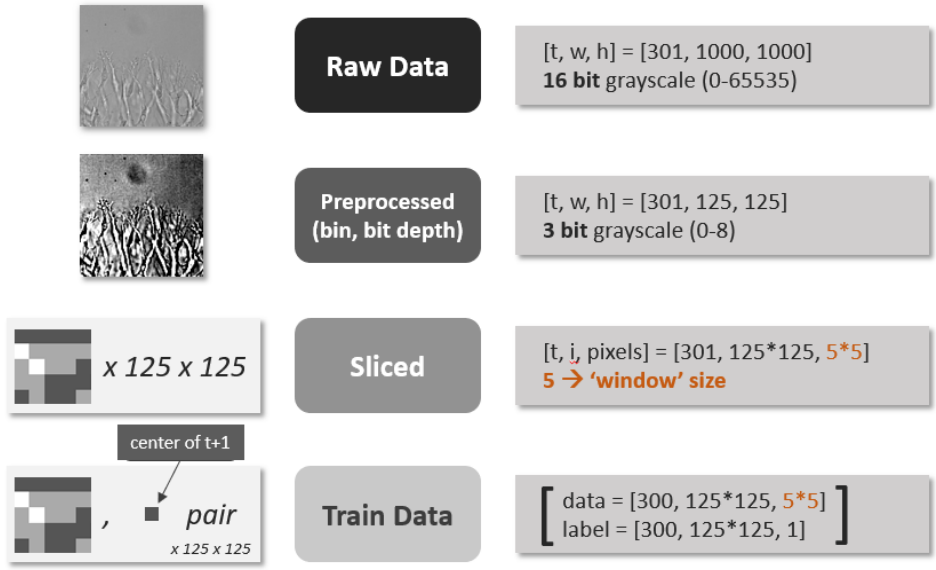


Figure 2.1. Data reshaping. Live image sequence data is reshaped to trainable data during ICA process.

In this study, we utilized Tensorflow™ [17], a Python based machine learning library presented by Google Inc. The software first reshapes the given image sequence into the trainable data set which is discussed in 2.2.2. Then the samples are given as inputs to the recurrent neural network model. Here, the image bit range is downsampled 4 bit grayscale image. In this study, we applied LSTM model [14] with 1024 hidden layer size.

## 2.3. Discussion

### 2.3.1. Tuning Hyperparameters

There were several hyperparameters used in this study. First, window size refers the size of the neighborhood area including target cell to apply CA rules. If we enlarge the window size, ICA will include more large area to learn the CA rules. This might conclude into more accurate rule prediction, however, possible number of CA rules also increases exponentially. This is because number of possible rules is described as

$$\#\ \text{of possible CA rules} = \text{possible cell status}\text{\^{}}(\text{window size}\text{\^{}}2),$$

which will infeasible to learn via given number of samples. In this study, we used the window size 5 for most of the image sequences and modified the image size (i.e. width * height) to fit the framewise image feature change will included within the given window size. One should modify the image size rather than the size of the window in order not to increase the size of the possible set of rules which will be increase computational cost exponentially.

Second, we should consider the size of the hidden layer. So far, there are no perfect way to define the size of the hidden layer, and one should decide how many layers will be feasible to learn the ICA rules via trial and error. In this study, we used the LSTM hidden size as 1024, however, this might be vary according to given computing power or structure, size of the given sample image sequence.

Third, number of dropout probability [18] should be in consider to design the ICA leaning model. This number also affects the accuracy of the prediction. In this study, we used 0.7~0.9 dropout probability, Generally, higher dropout probability will end with more accurate prediction model.

Other hyperparameters such as learning rate and exponential learning rate decaying parameter also could be considered in order to tune the learning speed and final prediction quality.

# Chapter 3.

# Learnability of ICA

## 3.1. Introduction

In this chapter, we will discuss the learnability of the inverse cellular automata model. To proof the concept of the inverse cellular automata model, we decided to introduce ICA to existing popular CA model-driven image sequences. In this study, we introduced to different 2-D CA rules that represents the cellular automata image sequence, ‘Conway’s Game of Life’ [8], and ‘Vichniac Vote’ [19].

## 3.2. Methods

### 3.2.1. Existing CA image sequence

We picked two different cellular automata rule sets to mimic the actual biological image sequence. Conway's Game of Life (GOL) and Vichniac Vote were chosen among the many existing rules since they are non-linear, and well-known as successfully mimics the biological complexity. GOL rules of life are:

- Any live cell with fewer than two live neighbours dies, as if caused by underpopulation.
- Any live cell with two or three live neighbours lives on to the next generation.
- Any live cell with more than three live neighbours dies, as if by overpopulation.
- Any dead cell with exactly three live neighbours becomes a live cell, as if by reproduction.

These simple four rules generate very complex pattern and produces many 'behaviors' that resamples complex cellular system.

Vichniac Vote is another CA that mimics the pattern of animal print such as cow. The rule for Vichniac Vote are:

- If the cell is in majority, it remains the previous status.
- If the cell is in minority, it change the status to the majority.

### 3.2.2. Software

The software for this study was built with Java-based image programming library. Initially, the software generates randomized pattern, and it generates next generation of the pixel images via given cellular automata rules. And the software captures the image sequence. Finally, produced image sequence is passed into the ICA learning part described in Chapter 2. When the given learning step ends, the software automatically generates the ICA based image sequence.

For GOL, we used 5000 epoch for 1000 batch rule samples, and 0.4 learning rate with 0.9 dropout probability. For Vichniac Vote, we used 2000 epoch with 300 batch rule samples with 0.4 learning rate and 0.9 dropout probability.

## 3.3. Results & Discussion

[Figure 3.1] shows the ICA based CA rule learning and its results. With given 100 * 100 * 120 frames of GOL image sequence, ICA generates 100 * 100 * 119 rule samples to be applied to the recurrent neural network. It successfully regenerates the rules of the GOL and follows perfectly without any loss.

However, the ICA based rule learning was highly dependent on several hyperparameters. In this study, the dropout probability was the one of the key factors to provide good prediction quality. With 0.5 dropout probability, ICA failed to reproduce GOL rules perfectly. [Figure 3.2, 3.3]

We also found the border condition should be considered during the rule learning. When the target cell is on the edge of the image, the window for the rule learning covers the outside of the original image. In this case, the ICA also deals the rules for the outside of the image frame, which is infeasible to most cases. For the continuous rule learning, we applied zero-padding for most of the ICA learnings as most Convolutional Neural Network does for the deeper image kernel learning. However, if ICA also considers the border with zero-padded, the rule learning will

end with failure since the zeros in the border might not be the proper value for the rules for next status of the target cell. [Figure 3.4] shows the case, where there are many zero values were padded to the Vichniac Vote original rule samples. In this case, the result showed well-trained pattern reproduction as [Figure 3.4] shows, however, there are some errors at the edge of the image frame. Thus, we learned that to apply ICA to general case, one should consider the value of the neighborhood target cells for the edge of the given image sample.

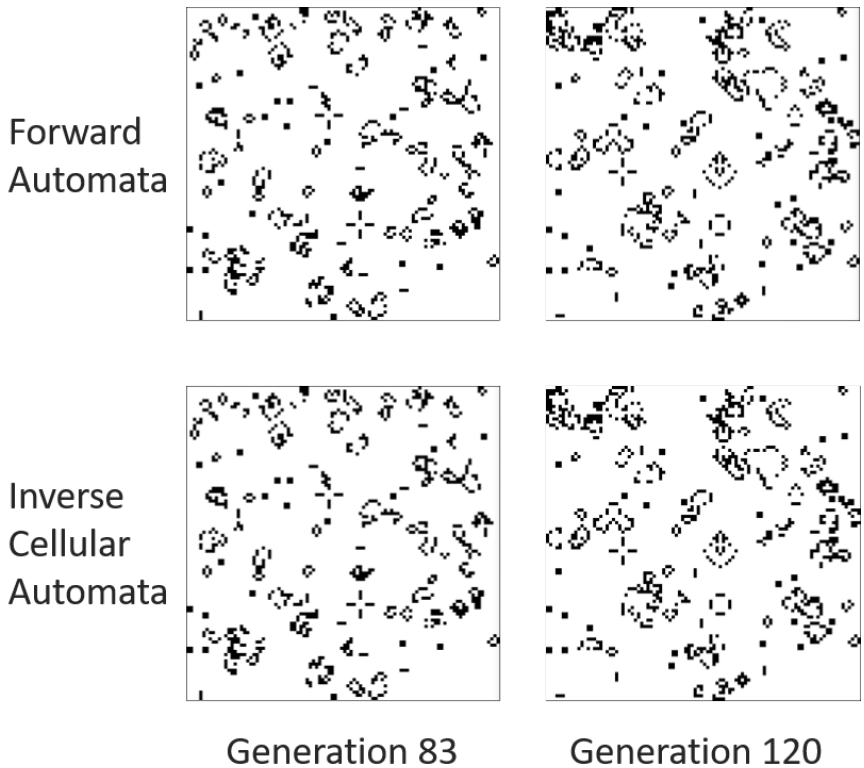


Figure 3.1. Result of the ICA based image prediction (Game of Life).

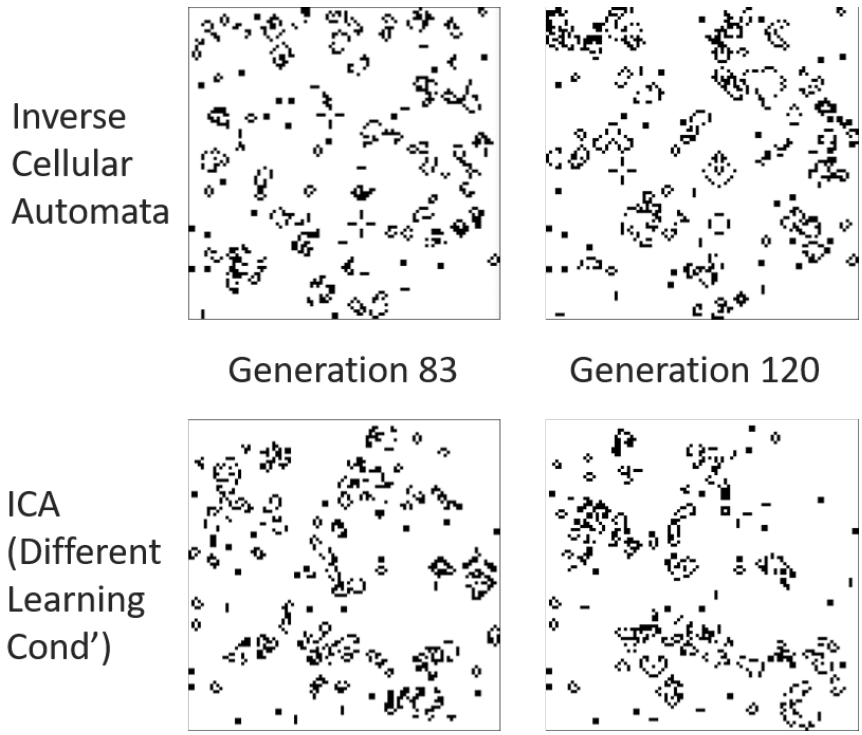


Figure 3.2. Result of the ICA based image prediction (Game of Life). With different learning Condition.

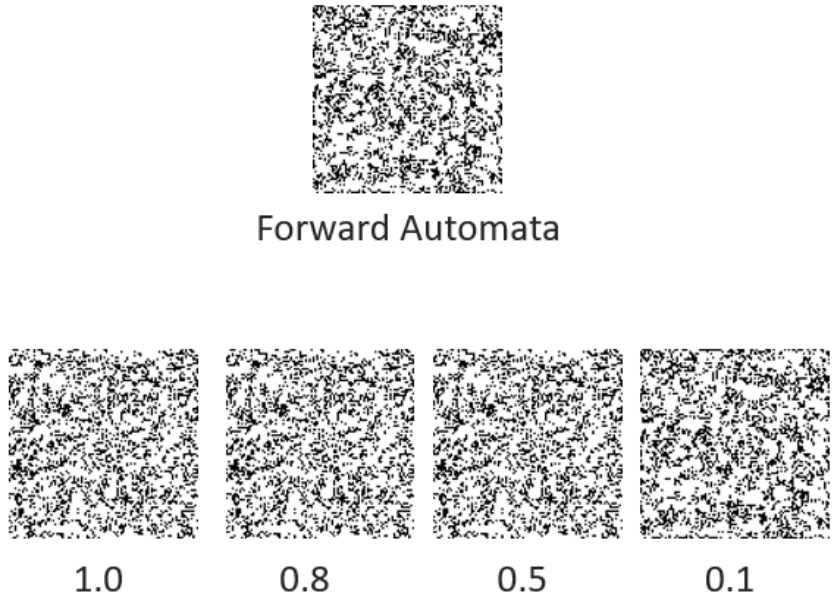


Figure 3.3. Result of the ICA based image prediction (Game of Life). With different dropout probability.

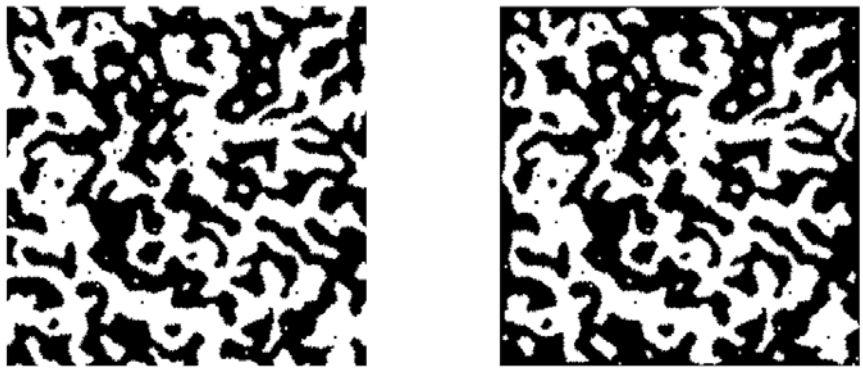


Figure 3.4. Result of the ICA based image prediction (Vichniac Vote).

# Chapter 4.

# Biological Image Sequence Application

## 4.1. Introduction

To apply the ICA-based image prediction in biological application, we considered several image sequences that can be show the possibility of the ICA properly. In this study, we applied the time-lapsed image sequence of the Physarum polycephalum.

Physarum polycephalum, also known as ‘slime mold’, is a kind of eukaryotic organisms. They are usually behaves in multicellular system, and shares information via internal pulse with neighborhood cells. Recently slime mold also received attention because of its possibility of solving ‘Salesmen Problem’ [20]. Because of these multicellular, information exchange features of Physarum polycephalum, we chose the specimen as the actual biological imaging prediction application.

## 4.2. Results and Discussions

Image sequence were gathered from [21]. Original time-lapsed video was preprocessed into 4 bit grayscale and we enhanced contrast of the image sequence. To avoid the edge learning problem discussed in Section 3.3, we specified the learning ROI (x:51, y:136, width:120, height:120 pixel) into small focused area of the original video, which the Polycepharum starts growth near the center of the video.

[Figure 4.1-3] shows the predicted Polycepharum growth process via ICA. Since the trained area was restricted to the left-downward direction, predicted ICA also directed to the left-downwarded. Since we downscaled the bit depth from 8 bit RGB into 4 bit grayscale, final image prediction also has the 4 bit image. However, we can find that predicted image sequence (Right side of Figure 4.1-3) follows the growth of the original time-lapsed image sequence of Polycepharum (Left down corner of Figure 4.1-3). The growth shows the complex growth direction biased to the left-downward and chaotic branching pattern as the original Polycepharum image sequence shows. This shows that the ICA can successfully learned the growth pattern of the Polycepharum from the image sequence.

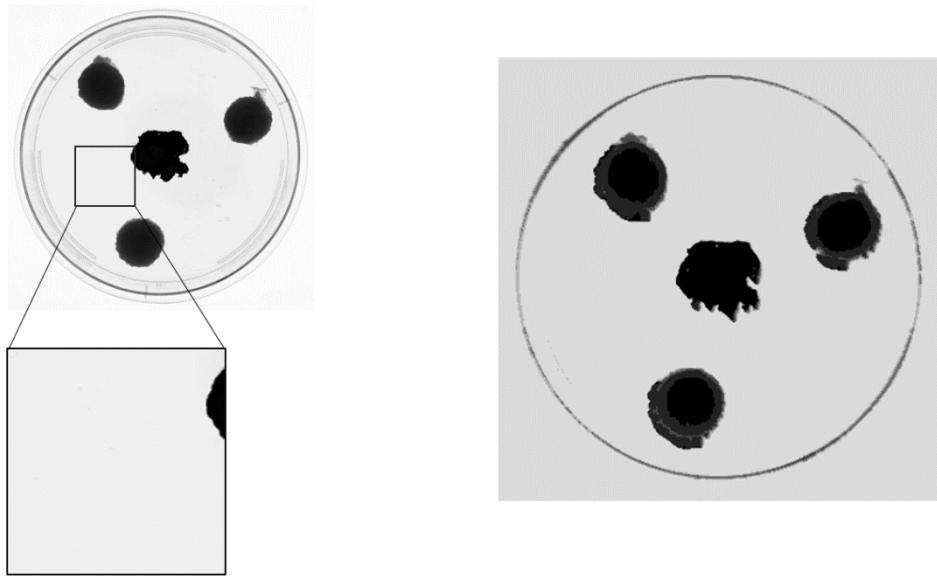

Figure 4.1. ICA trained Polycepharum process. Frame 10.

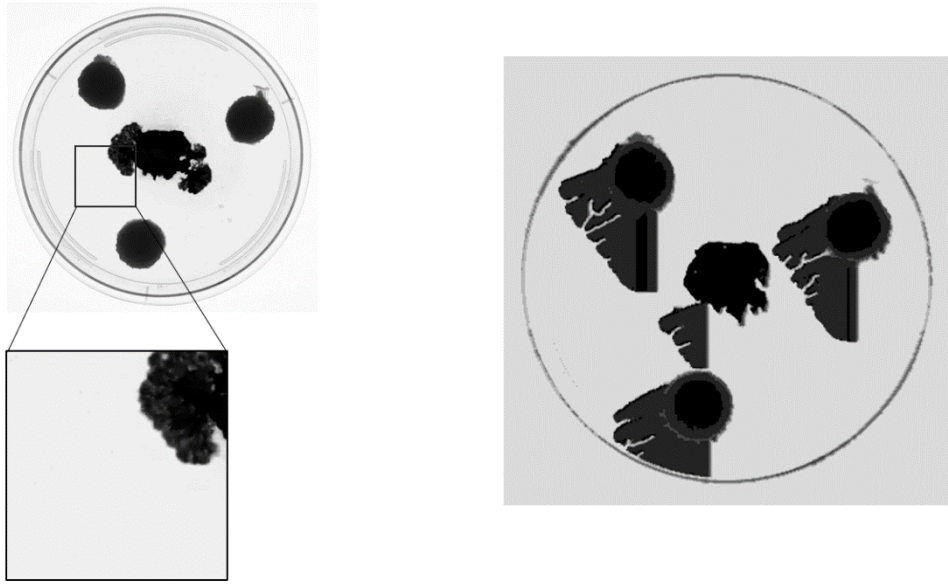

Figure 4.2. ICA trained Polycepharum process. Frame 50.

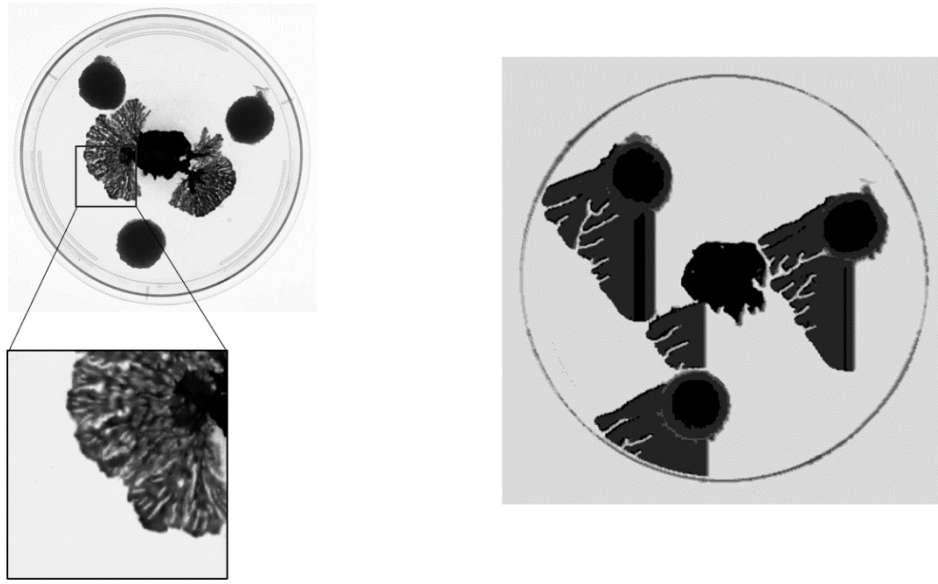

Figure 4.2. ICA trained Polycepharum process. Frame 100.

# Chapter 5.

# Conclusion

Investigating biological complex pattern requires a lot of effort. Researchers first need to observe the pattern intuitively and analyze that what factors would make the complexity one by one. This approach is called reductionism. In reductionism, we investigate the factors that composes the complexity individually, and hopes that once we understand all the individual factors than we can understand the complexity. However, this approach might lead to fail when we want to predict the complexity and reproduce them on the lab. If we do not understand the individual factors properly, the simulation could end with failure. Or, some small unbalanced weighting factor could also lead the simulation failure.

Instead, we can approach the problem as holistic manner. When we deal with the problem as holistic manner, we remains the internal working mechanisms into the black box, and focuses on the prediction of the emergent factors. We claim that the ICA would be the one of the great approaches to achieve the holistic investigation and prediction of the biological complexity. Although we do not fully understand the

working patterns individually, we can leave the internal details to the neural network based machine learning, and focuses on the prediction of the complex emergent pattern itself.

# Bibliography


[1] Pang, B., L. Lee, and S. Vaithyanathan. Thumbs up?: sentiment classification using machine learning techniques. in Proceedings of the ACL-02 conference on Empirical methods in natural language processing-Volume 10. Association for Computational Linguistics,2002.

[2] Nasrabadi, N.M., Pattern recognition and machine learning. Journal of electronic imaging. 16(4): p. 049901,2007.

[3] LeCun, Y., Y. Bengio, and G. Hinton, Deep learning. nature. 521(7553): p. 436,2015.

[4] Schmidhuber, J., Deep learning in neural networks: An overview. Neural networks. 61: p. 85-117,2015.

[5] Wolfram, S., Cellular automata as models of complexity. Nature. 311(5985): p. 419,1984.

[6] Von Neumann, J., The general and logical theory of automata. Cerebral mechanisms in behavior. 1(41): p. 1-2,1951.

[7] Schiff, J.L., Cellular automata: a discrete view of the world. Vol. 45. John Wiley & Sons, 2011.

[8] Conway, J., The game of life. Scientific American. 223(4): p. 4,1970.

[9] Wolfram, S., A NEW KINDS OF SCIENCE, Wolfram Media. Inc,2002.

[10] Cheng, G., P. Markenscoff, and K. Zygourakis, A 3D hybrid model for tissue growth: the interplay between cell population and mass transport dynamics. Biophysical journal. 97(2): p. 401-414,2009.

[11] Qi, A.-S., et al., A cellular automaton model of cancerous growth. Journal of theoretical biology. 161(1): p. 1-12,1993.

[12] Lee, Y., et al., A cellular automaton model for the proliferation of migrating contact-inhibited cells. Biophysical journal. 69(4): p. 1284-1298,1995.

[13] Mishra, A., et al. A cellular automata based semi-automatic algorithm for segmentation of choroidal blood vessels from ultrahigh resolution optical coherence images of rat retina. in Electrical and Computer Engineering (CCECE), 2010 23rd Canadian Conference on. IEEE,2010.

[14] Hochreiter, S. and J. Schmidhuber, Long short-term memory. Neural computation.

9(8): p. 1735-1780,1997.

[15] Mikolov, T., et al. Recurrent neural network based language model. in Eleventh Annual Conference of the International Speech Communication Association.2010.

[16] Zaremba, W., I. Sutskever, and O. Vinyals, Recurrent neural network regularization. arXiv preprint arXiv:1409.2329,2014.

[17] Abadi, M., et al. TensorFlow: A System for Large-Scale Machine Learning. in OSDI.2016.

[18] Gal, Y. and Z. Ghahramani. A theoretically grounded application of dropout in recurrent neural networks. in Advances in neural information processing systems.2016.

[19] Vichniac, G.Y., Simulating physics with cellular automata. Physica D: Nonlinear Phenomena. 10(1-2): p. 96-116,1984.

[20] Watanabe, S., et al., Traffic optimization in railroad networks using an algorithm mimicking an amoeba-like organism, Physarum plasmodium. Biosystems. 105(3): p. 225-232,2011.

[21] SwarmLabNJIT. The 3 stages of Physarum polycephalum. [cited 2017 Sep. 17]; Available from: https://www.youtube.com/watch?v=qCOv0F7kPlU,2016

국문초록

# 재귀 신경망이 적용된 역 셀룰러 오토마타 학습법을 활용한 세포 분지(分枝) 예측

서울대학교 대학원
기계항공공학부
윤 경 원

생물학적 과정은 매우 복잡해서, 그 과정을 쉽게 예측하기 어렵다. 이미지를 기반으로 한 생물학 프로세스 예측에 있어서 가장 큰 어려움은 패턴을 예측하기 위한 충분히 많은 시각 데이터를 실제 생물학 실험을 통해 많이 획득하기가 어렵다는 데에 있다. 이 논문에서는, 세포 분지 등의 생물학 프로세스 예측을 위해서 역 셀룰러 오토마타법이라는 새로운 방법을 제안한다. 이 방법은 생물학적인 프로세스를 촬영한 이미지 시퀀스가 $t$에서 $t+1$로 진행되는 셀룰러 오토마타 시스템이라는 것을 가정한다. 이 가정을 바탕으로, 하나의 이미지 시퀀스는 폭*높이*프레임 수 만큼의 셀룰러 오토마타법 규칙 집합을 추출할 수 있고, 이 충분한 수의 규칙 집합을 기존의 재귀 신경망 학습법에 적용할 수 있다. 본 연구에서 제안하는 방법을 통해 기존의 생물학 이미지 예측법에서 샘플의 수가 부족했던 문제를 크게 개선할 수 있다. 연구에서는 우선, 기존의 셀룰러 오토마타법을 되돌아보고, 역 셀룰러 오토마타법의 개념에 대해 설명한 뒤, 기법의 학습 가능성을 기존에 존재하는 셀룰러 오토마타법 규칙을 통해 만든 이미지를 통해 검증한다. 마지막으로 실제 생물학 이미지 시퀀스를 학습하여 역 셀룰러 오토마타법을 통해 학습된 결과를 소개한다.

**주요어** : 셀룰러 오토마타법, 생물학 이미징, 이미지 추정, 인공 신경망, 재귀 신경망
**학 번**: 2012-30924